University at Buffalo
The State University of New York

Language Error in Aviation Maintenance

Year 3 Interim Report: Asia

C. G. Drury and J. Ma

January 2005

Prepared for:

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405
Contract #2002-G-025
Dr. William Krebs

TABLE OF CONTENTS

List of Figures

List of Tables

EXECUTIVE SUMMARY

This report extended the University at Buffalo/Federal Aviation Administration study of language errors in aviation maintenance to one world region: Asia. Based on earlier studies in the USA and UK, plus analyses of a number of databases, seven scenarios for language error were developed, and the current study assessed the incidence of these scenarios in site visits to nine sites covering 254 participants in China, Hong Kong and Taiwan. A second study on the same sample was a direct test of the effectiveness of four interventions. Finally, focus groups were used at each site to explore strategies for mitigating language errors. The level of cooperation with the data collection team at all sites was outstanding.

The scenario incidence study confirmed the scenarios as valid and no new ones were added from the focus groups. The typical picture for a language error-prone activity was one with complex task and complex instructions, poorly designed document, users with low ability in English and low familiarity with the task to be performed and with time pressure to complete the task.

The interventions experiment used a baseline condition of English documents, and then added translation (including the test form), a glossary, a bilingual coach, and a combination of these last two conditions. We used two levels of workcard difficulty, each with and without Simplified English. While there were some differences between regions, differences between interventions were consistent across regions. The translation intervention had some effect, although mainly on the times and our performance measure, rather than on accuracy *per se*. If this indeed reflects practice, then maintenance personnel appear to slow down when they find language difficult, rather than making more errors at a constant speed.

Several practical interventions emerged from all three data collection methods. Design of work documentation is the primary way to reduce written language errors. Good design practice will help and translation, if performed carefully, is a viable option. Individual ability of Aviation Maintenance Technicians (AMTs), inspectors, managers and engineers in written and verbal English communication can be improved by training and controlled practice. The organizational environment should recognize the deleterious effects of time pressure on errors, and also recognize the symptoms of imperfect communication when it occurs. The organization also needs to plan work assignments to allow AMTs to become more familiar with particular tasks, and to use controlled implementation of English in shift turnover documents and non-routine repair documents to provide planned English practice for all personnel.

Our next task is to repeat this experiment in other continents. The current plan is to visit locations in Central and South America and Europe in Spring 2005.

1.0 INTRODUCTION

In 2001, the Federal Aviation Administration (FAA) raised many issues concerning the outsourcing of maintenance to foreign repair stations in considering changes to domestic and foreign Federal Air Regulations, recommending that:

> *"The FAA should establish a method for determining whether language barriers result in maintenance deficiencies."*

This project is a direct response to these concerns that non-native English speakers, in repair stations in the USA and abroad, may be prone to an increased error rate that could potentially affect airworthiness. The documentation for repair provided by an English speaking airline is always in English, and this documentation must be used to govern all maintenance tasks, despite a potentially large proportion of mechanics who do not use English as a native language. This report follows our 2004 Human Factors and Ergonomics Society paper (Drury and Ma, 2004)[1] and describes data collection trips to Asia using a methodology for quantifying the effectiveness of possible countermeasures to language errors.

As noted in our 2004 paper, this project developed seven scenarios of language error based on visits to sites in the USA and the UK; it also provided a model for these unique communication errors based on the communications literature and an analysis of several databases (e.g., NASA/ASRS). Many references to communication theories and studies of outsourcing were given in Drury and Ma (2003[2]; 2004[1]) and will not be repeated here.

The seven scenarios found were:

Scenario 1: "The Mechanic (Aircraft Maintenance Technician, AMT) or Inspector was not able to communicate verbally to the level required for adequate performance."
Scenario 2: "The Mechanic (AMT) or Inspector and the person to whom they were speaking did not realize that the other had limited English ability."
Scenario 3: "Native English speakers with different regional accents did not understand each others' communications."
Scenario 4: "The Mechanic (AMT) or Inspector did not understand a safety announcement over the Public Address (PA) system."
Scenario 5: "The Mechanic (AMT) or Inspector did not fully understand a safety placard."
Scenario 6: "The Mechanic (AMT) or Inspector did not fully understand documentation in English, for example a Workcard or a Manual."
Scenario 7: "The Mechanic (AMT) or Inspector did not fully understand a document translated from another language into their native language."

In our work, we have been visiting sites worldwide to measure the frequency of these scenarios, and evaluating the effectiveness of countermeasures. An intervention experiment has been designed and tested using a sample of 254 maintenance personnel from countries in Asia. In addition, data on reported frequency of these scenarios and factors associated with their occurrence was collected on the same sample.

A survey conducted by a major manufacturer showed that English language skill varied (as expected) by world region, and that not all sites with lower language skills translated documents into the native language. Our analysis of the survey data reported earlier found that two strategies used to reduce the potential for language errors were (a) translation into the native language, and (b) conducting face-to-face meetings in the native language. However, only about 17% of airlines in the region that most often used translation (Asia) actually translated maintenance documents into the native languages. Even among the group of 8 airlines who reported the lowest English speaking ability, only 2 modified the English documents in any way. Other strategies of intervention found in our site visits included having a bilingual English/native language speaker (e.g., lead, engineer) assist the mechanic with the English documentation, and/or providing a glossary of key words between the native language and English. Finally, our own earlier research into the artificial maintenance language called European Association of Aerospace Industries (AECMA) Simplified English (e.g., Chervak, Drury and Ouellette, 1996[3]) had shown it to be an effective error reduction technique, particularly for non-native English speakers and for complex work documents. Thus, we planned to compare four potential language error reduction interventions:

1. The translation of a document into AECMA Simplified English
2. The provision of a Glossary
3. The provision of a bilingual coach
4. The translation of a document and all related materials into a native language

Some of these methods can be combined, for example the provision of both a Glossary and a bilingual coach, or the addition of AECMA Simplified English to all conditions except for translation into the native language. Finally, for comparison, a baseline condition, no intervention, was required. This paper describes briefly the first two experiments conducted within this framework, and the main data collection in one region, Asia.

2.0 METHODOLOGY

Three aspects of interest formed the basis for our data collection efforts, designed specifically to answer FAA questions about the nature and frequency of language errors and possible interventions to reduce these errors.

First, typical demographic measures were collected for each participant: Age, Gender, Job Category and Years as an Aviation Maintenance Technician (AMT). In addition we collected language data to characterize the population of AMTs and also to provide potential covariates for our analyses of intervention effectiveness. These were Years Studying English, and a measure of reading grade level from the Accuracy Levels Test. Second, a questionnaire was given for each scenario asking whether the respondent had encountered that scenario, how long ago, and what were the factors associated with the scenario. Third, the set of interventions noted above were tested using a workcard comprehension measure to find their effectiveness. Finally, one or more focus groups were held at each site to better understand the way in which potential language errors were handled in their organization.

2.1 Measures

Demographic data were collected as noted above. The Accuracy Levels Test (Carver, 1987[4]) used a total of 100 words with a forced synonym choice among three alternatives (10 minute maximum), and produced on the scale of reading grade level normed on US public schools. It has been validated against more detailed measures of reading level (Chervak, Drury, Ouellette, 1996[3]).

For each of the seven scenarios the incidence questionnaire first asked whether each had ever been encountered. This was the primary incidence measure, i.e. percentage incidence of each. To get more detail on frequency, respondents were asked whether the scenario occurred in the past week, month, year or longer. We also asked how many months or years, but the data were not always given in a quantitative manner, so an estimate of the mean time since previous occurrence was derived from the week/month/year data. Also for each scenario, participants were asked to check the factors associated with increased likelihood of the error occurring (9 factors), with mitigating each error (10 factors) and with the discovery of each error (6 factors). The factors came from our previous analyses of databases of errors and focus groups used to derive the scenarios (Drury and Ma, 2003[5]).

To test for how potential documentation errors can be reduced, we measured the effectiveness of document comprehension. In the study, a single workcard was given to participants with a 10-item questionnaire to test comprehension. The methodology had been validated in our previous research (e.g., Chervak, et al., 1996[3]; Drury, Wenner and Kritkausky, 1999[6]). The comprehension score was measured by the number of correct responses, with time taken to complete the questionnaire as an additional measure. In addition, the workcard was rated by the participant on the fifteen scales originally developed by Patel et al (1994[7]).

2.2 Workcards

We selected two workcards, one “easy” and one “difficult,” from four workcards used in our previous research (Drury, Wenner and Kritkausky, 1999[6]), because it had already been found that task difficulty affected the effectiveness of one strategy, Simplified English. As was expected, the use of Simplified English had a larger effect on more complex workcards (Chervak and Drury, 2003[8]). The complexity of these workcards was evaluated by Boeing computational linguists and University of Washington technical communications researchers considering word count, words per sentence, percentage passive voice, and the Flesch-Kincaid reading score. The cards differed on all measures. Note that both cards were comparatively well-written, certainly compared to earlier workcards tested by Chervak et al (1996[3]).

Both of the workcards were then prepared in the AECMA Simplified English versions, which were also critiqued by experts from Boeing, the University of Washington, and the American Institute of Aeronautics and Astronautics (AIAA) Simplified English Committee.

2.3 Pre-Test Design

To test the design and materials, two pilot studies were conducted, one using 15 English-speaking maintenance personnel from sites in the USA and the UK, and the other using 40 Native Chinese speaking engineering graduate students at the University at Buffalo, SUNY. These tests successfully proved the evaluation methodology, and eliminated one condition (glossary plus bilingual coach) as participants did not make use of both. Full details were given in our 2004 Human Factors and Ergonomics Society paper (Drury and Ma, 2004[1]).

2.4 Experimental Design for Comprehension Test

A fully nested (between subjects) $2 \times 2 \times 4$ design was used with factors as follows:

Workcard Complexity: 2 levels	- Simple - Complex
Workcard Language: 2 levels	- Simplified English - Not Simplified English
Language Interaction: 4 levels	- No intervention (English) - English with glossary - English with coach - Full Chinese translation

We originally intended to use an additional intervention using glossary plus coaching to form a 2 glossary × 2 coaching sub-design. However, in both our Chinese Engineering Graduate sample and early tests at maintenance, repair and overhaul (MRO) sites, it became obvious that very few participants actually used these job aids. Thus, the combined intervention of glossary and coaching was dropped from the study. Additionally, the Chinese translation of each workcard was only performed once, whether for Simplified English or not, so no difference was expected between Workcard Language for that intervention.

2.5 Choice of Participants and Sites

Note: the political status of the three "countries" selected, China, Hong Kong and Taiwan is complex, so they will be referred to in this report as "Areas" to avoid the impression that they are, or are not, parts of the same country. There are several reasons to collect data from MROs located in Asia, especially China, Taiwan and Hong Kong. First, in our analysis of the manufacturer's survey data, we found that about 30% of users in Asia had a very limited English speaking ability, another 40% were able to conduct simple conversations; about 40% of the users were able to work effectively with only written maintenance/inspection related documents, and another 15% had very little English reading ability. Compared with North America and Europe, Asia has a much smaller base of English-using mechanics. Second, the Asia-Pacific region is poised to be one of the strongest growth engines for the foreseeable future for the maintenance, repair and overhaul industry (*Overhaul & Maintenance*, 2002[9]). U.S. and European airlines continue to ship wide-body aircraft to East Asia to take advantage of low labor costs. Almost half of the top ten Asian MROs are located in China. According to *Aviation Week & Space Technology*, "the Civil Aviation Administration of China (CAAC) is confident that despite the downturn in the global airline industry, more maintenance, repair and overhaul (MRO) joint venture companies will be set up with Chinese airlines within the next two years" (Dennis, 2002[10]).

In addition, from our initial collections of patterns of language errors in English-speaking countries (USA, UK) and our analysis of the language database, it was apparent that to collect a broad range of data, Asia would be the appropriate region. Asia is a major growth region for third party MRO work, and was also likely to provide a wide range of managerial practices for handling language differences. We have already observed the use of English coaching by a more senior person, e.g., lead, foreman, engineer. Also, from the airline survey, we learned that some organizations translate documents into the native language of the employees. Finally, we have seen glossaries of English/native language words pertaining to aviation maintenance. All three are managerial practice we wish to examine through experimentation.

2.6 Preparation of the Data Collection Packet for Asia

Contacts in several Chinese-speaking countries were helpful in gaining access to MROs. The translation process took place in two steps. A native Chinese research assistant (nine years as an engineering major), who is very familiar with the workcards and fluent in English, took a lead in translating the packet. A large number of technical and language references were consulted. The principal investigator and other domain experts (e.g., native Chinese mechanical engineers in the Department of Aerospace and Mechanical Engineering at the University at Buffalo, SUNY) were consulted on the technical details (e.g., lockwire). Then both translated and original packets of data collection material were submitted to a retired professor (also fluent in English) from the Department of Avionics, Civil Aviation University of China (CAUC) for review. The translated material included the four workcards (for the full translation condition), the comprehension questions, the workcard ratings, the demographic information form, the informed consent form and the questionnaire on frequency and causality of language errors.

We developed an English/Chinese glossary for each workcard. We had two native English speaking engineering graduate students and two native Chinese speaking engineering graduate students read through all the workcards and circle all the words/phrases/sentences they did not comprehend, or even those about which they were slightly unsure. We built up this glossary to be as comprehensive as possible, including nouns, verbs, adjectives, abbreviations, etc.

For data collection where traditional Chinese was used (i.e., Taiwan), all forms were checked for correct current usage of traditional Chinese characters by two bilingual Chinese/English engineers with good knowledge of both human factors and aviation maintenance.

We also prepared for data collection in an Asian country with a different language, but the MROs cancelled data collection prior to our visit.

2.7 Data Collection Process

At each MRO site, an initial meeting with management was used to explain verbally the objectives and conduct of the study, as a supplement to our earlier written communications. At this meeting, we also discussed the type of work at the site, the range of customers served and the importance of language issues and errors. Agreement was reached on the types of participants of most use to us, e.g. AMTs, engineers, QA personnel, managers. The company then scheduled multiple participants at approximately 75 minute intervals.

Groups of participants were nominally of six people, but groups with 2-10 were encountered. Each group of participants was welcomed, and the general objective of the data collection, i.e. to understand language errors and how to reduce them, was communicated. After obtaining Informed Consent and completing demographic questions, the participants all started the timed intervention evaluation (workcard comprehension test) at the same time. The participants were given one of the four workcards and its associated comprehension questions. They were timed, but instructions emphasized accuracy. When this had been completed, each participant was given the rating form. The participants who finished both of these rapidly were given the seven-scenario frequency/causality questionnaire to allow slower participants to catch up. All were then given the Accuracy Levels Test, starting the 10 minute timed test at the same time. If time remained in the 75 minute session, the participants who had not completed the incidence questionnaire were given that. If there would not be time, remaining participants were asked to take their questionnaires back to their workplace and return the completed questionnaires later. The participants were individually thanked for their participation and given a small gift: a plastic water bottle marked with "UB: Official Experimental Participant."

The participants were scheduled to be tested in groups with the same intervention, as far as possible. However, at times too few participants arrived so that mixed groups were sometimes tested. The participants were told that not all people in the experimental room were getting the same workcard, or the same intervention condition. On a couple of occasions, a participant did not even attempt the task in one of the first three intervention conditions because they did not read English. In these few cases, the response was noted and the participant was given the equivalent full Chinese translation condition. We could later count this participant as scoring zero on the comprehension test if required.

The participants were assigned to the workcard complexity and workcard language conditions in rotation. As they were assigned to the experiment by their manager, no unwanted volunteer bias from this procedure was expected. The participants were volunteers in the experiment, but only after they had been assigned to attend by their managers.

A total of 13 focus groups, each of 6-15 engineers, quality personnel, AMTs and managers, were conducted across the sites. Discussions were wide-ranging in English and Chinese, led by one or more of the experimenters posing questions about language errors, communication problems, their causal factors and how such errors are mitigated by the organization and its people.

3.0 RESULTS

A general description of the characteristics of each site is presented in Appendix Table 1. The data were collected from written sources, managers and focus group discussions. A primary result of this data collection was the finding that all of the sites in China used a mixture of English and translated Chinese documentation, while in Hong Kong and

Taiwan only English documentation was used.

3.1 Demographics

For each participant we recorded their Gender, Age, Years as an AMT, Years Learning English and Reading Level as given by the Accuracy Levels Test. As it was most unlikely that these demographics would remain constant across the three areas (China, Hong Kong, Taiwan), one-way ANOVAs were conducted of each demographic, except for the categorical variable of Gender that was tested using Chi-Square. All comparisons gave significant differences by Area, as shown in Table 1.

Note that China has more females represented, in a generally younger, less experienced sample with lower English exposure and reading ability. For Years as AMT and Years Learning English, all three areas were significantly different, with Taiwan falling between the low value of China and the high value of Hong Kong.

There were no gender differences among the demographic variables using a 2 factor GLM ANOVA of Area and Gender, except for Years as AMT where females (6.1 years) were less experienced than males (11.6 years) with $F(1, 247) = 6.6$, $p = 0.011$.

	China	Hong Kong	Taiwan	UK/ USA	Test Result	Significance
Number Tested	175	25	54	15		
Percent Female	25%*	4%	4%	0	$\chi^2(2) = 15.84$	$p < 0.001$
Age	33.5*	42.9	40.5		$F(2,250) = 34.7$	$p < 0.001$
Years as AMT	8.6*	18.4*	13.6*		$F(2,250) = 21.9$	$p < 0.001$
Years Learning English	20.1*	35.6*	27.1*	-	$F(2,243) = 79.9$	$p < 0.001$
Reading Level	4.9*	6.6	5.8	14.1	F(2,253 = 7.9	$p < 0.001$

Table 1. Demographics of the three areas, with mean values and test results. Note that * signifies a mean value different from the others at p<0.05 on the *post hoc* Tukey test, or Standardized Residuals test for Chi-square.

3.2 Incidence Questionnaire

In addition to the evaluation of the interventions, we used a questionnaire to determine the relative incidence of the seven scenarios developed earlier. A number of measures of incidence were used, including estimates of the time since last occurrence. The first analysis was of the overall response to “Have you ever encountered an error of this type?” A two-factor GLM ANOVA (Scenario x Area) of whether or not each scenario was reported resulted in significance for Scenario $F(6, 1722) = 28.2$, $p < 0.001$, for Area $F(2, 1722) = 5.3$, $p = 0.005$, and for their interaction $F(12, 1722) = 2.7$, $p = 0.002$.

Overall, the scenarios group into three sets using the Tukey post-hoc test. The most frequent set (Scenarios 6 and 1) refer to the AMT not understanding written (6) or verbal (1) instructions. The next set of three (Scenarios 7, 2 and 3) refer to poor translation of documents (7) often from English by aircraft manufacturers for whom English is not the native language, not realizing that the AMT did not understand (2) or difficulties with regional accents (3). The least frequent set (Scenarios 5 and 4) consisted of relatively rare forms of communication, placards (5) and the PA system (4).

The three areas did not produce clear cut results in post-hoc tests. China reported a higher incidence (42%) than Hong Kong (32%), but was not different from Taiwan (39%). Also, Taiwan was not different from Hong Kong. Thus there is a hierarchy of incidence reporting, but only the extreme values are different from each other.

The interaction for incidence of each scenario is shown in Figure 1 for the three areas separately. Misunderstanding translations (Scenario 7) was highest in China and lowest in Taiwan, while the opposite ordering was found for misperceived language abilities (Scenario 2) and regional accents (Scenario 3). The first of these results is perhaps reflective of exposure, as Chinese sites used translation of parts of documents, which was not a strategy in the other two areas.

When the answers to the question “When was the most recent time you encountered on errors of this type?” were tabulated, it was possible to estimate the median time since the last occurrence of each scenario. A cumulative plot of probability of occurrence against time since last occurrence for each scenario was used to perform a linear interpolation of the median. The medians are shown for each scenario in Table 2 with the mean percentage reported from the previous analysis. As expected, the more frequently reported scenarios are the ones with the smallest median time since previous occurrence ($r = -0.817$, $p = 0.025$).

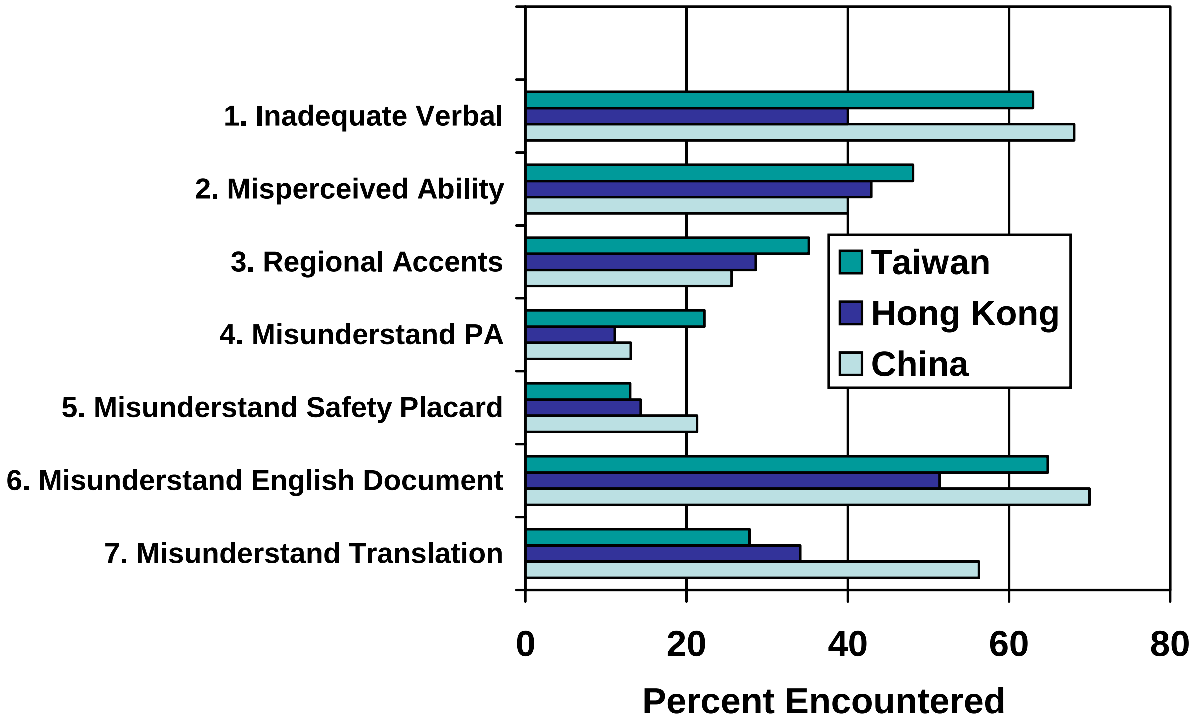

Figure 1. Relative frequency with which each of the seven scenarios was encountered.

Scenario	Median Weeks Since Previous Occurrence	Mean Percent Reported
1. Inadequate Verbal	16.0	39.4
2. Misperceived Ability	12.9	62.1
3. Regional Accents	21.5	16.2
4. Misunderstand PA	18.9	15.5
5. Misunderstand Safety Placard	18.0	29.8
6. Misunderstand English Document	9.2	43.7
7. Misunderstand Translation	12.0	57.0

Table 2. Median Weeks since Previous Occurrence and Mean Percent Reported for each Scenario

3.2.1 Error Factors

For the response to factors most associated with these scenarios, GLM ANOVA of the percentage encountering each incident by Factor was performed, with Area and Scenario as additional independent variables. All main effects and interactions except Scenario × Area were significant at $p < 0.02$ or better. Post hoc Tukey tests were performed at $p = 0.05$ to group the main effect levels of Factor. The responses divided into two groups, one group seen as highly related to the incident and one less related. Below these are given with their percentage reporting. These are:

Highest Related to Scenarios

The task is complex	35%
The task instructions are complex	41%
The mechanic (AMT) or inspector has inadequate written English ability	38%
The mechanic (AMT) or inspector has inadequate verbal English ability	36%
Time pressure makes the mechanic (AMT) or inspector hurry	33%

Lowest Related to Scenarios

The communication channel, e.g. radio or PA, interferes with good communication	13%
Time pressure prevents the mechanic (AMT) or inspector from asking other people for help	19%
The mechanic (AMT) or inspector reverts to their native language under stress	17%
The mechanic (AMT) or inspector is unwilling to expose their lack of English	18%

3.2.2 Prevention Factors

A similar analysis was performed for the ten factors potentially mitigating language errors. The GLM ANOVA gave significance at $p < 0.01$ for Factor, Area, and their interaction. As with causal factors, the results grouped into two:

Highest Related to Scenarios

The document is translated into the native language of the mechanic (AMT) or inspector	36%
The document uses terminology consistent with other documents	36%
The document follows good design practice	34%
The mechanic (AMT) or inspector uses the aircraft as a communication device, for example to show the area to be inspected	34%
The mechanic (AMT) or inspector if familiar with this particular job	45%

Lowest Related to Scenarios

The mechanic (AMT) or inspector has taken and passed a comprehension tests	21%
The mechanic (AMT) or inspector was certified for that specific job	22%
There is a translator available to help the mechanic (AMT) or inspector	22%
Jobs are assigned to the mechanic (AMT) or inspector to job based on English ability	23%
The mechanic (AMT) or inspector is teamed with a native English speaker to perform the job	19%

As with causal factors, the highest group included the physical changes, plus in this case job familiarity. The lowest group was mainly individual and social interventions.

3.2.3 Discovery Factors

Finally, an analysis of how errors are discovered was performed. Only Scenario, Factor, and the Factor × Area were significant (at $p < 0.02$). Again, there was a grouping of the Factors, this time into 3 groups:

Highest Related to Scenarios

The mechanic (AMT) or inspector asked for assistance or clarification.	56%

Medium Related to Scenarios

The mechanic (AMT) or inspector appeared perplexed	36%
The physical error resulting from the language error was detected.	31%

Lowest Related to Scenarios

The mechanic (AMT) or inspector agreed with everything that was said.	13%
The mechanic (AMT) or inspector did not understand inspector's questions at buy-back.	12%
The mechanic (AMT) or inspector closed access prematurely (i.e. before buyback)	6%

From these groupings, note that the least commonly found were either an unusual behavior, or events later in the maintenance/inspection process.

3.3 Intervention Effectiveness

This test used 254 participants from nine sites in China, Hong Kong and Taiwan. First, as in the pre-tests, there was a negative correlation between accuracy (fraction of correct responses) and time (overall time to complete the task) for the comprehension test ($r = -0.170$, $p = 0.007$). This was not as large as in the pre-tests, but still a significant speed/accuracy trade off. A third measure was created by dividing Accuracy by Time to give a combined overall Performance score.

Among the demographic variables, there were inter-correlations among the measures in Years (Age, Years as AMT, Years Learning English) as would be expected, but no significant correlations of these variables with Reading Level. Another way to express this is that a Factor Analysis (using a Varimax rotation) needed only two factors to explain 86.3% of the variance in these four measures, with the first factor loading above 0.85 on all the "Years" factors and the second loading only on Reading Level. From these analyses of individual characteristics, two relatively orthogonal measures were chosen as potential covariates in the performance analyses: Reading Level and Age.

There were moderate correlations of accuracy with Years as an AMT ($r = -0.231$, $p < 0.001$) and both accuracy and time with Reading Level ($r = 0.351$, $p < 0.001$; $r = -0.250$, $p < 0.001$ respectively).

Because the Simplified English factor was not a true factor for the intervention of Chinese translation, a separate set of analyses was performed with that intervention removed. These results will be noted as similar to of different from the main analyses as each is performed. As an example, all of the correlation results in the previous paragraph were mirrored in the "no Chinese Translation" analysis.

GLM ANOVAs were performed for each measure (Accuracy, Time, Accuracy/Time) as well as Log_e(Time) because that was found to be more normally distributed than Time. The factors tested were Intervention, Area, Workcard Difficulty and Simplified English, with the two covariates of Reading Level and Age. All main effects and two-way interactions were included, but not higher order interactions due to multiple co-linearity effects. Part of that was due to the fact that the Chinese Translation intervention could not be used in Hong Kong as the participants there would only use original English documentation. The Intervention x Area interaction was dropped from the analysis because of this missing cell.

The results of the ANOVAs are summarized in Table 3. Note that the use of AECMA Simplified English had no significant effect on any measures. Also, no interactions among any factors reached significance, simplifying the interpretation of results. The two covariates were highly significant in all analyses, this helping to reduce the error terms and so increase the power of the other tests.

To illustrate the predictive power of the covariates, Figure 2 (at end of report) shows the four plots of two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age). While they clearly show relationships, the variance is quite high for all four plots: performance in workcard comprehension is more than just good English ability and lower age.

From Table 3, it is obvious that most of the variation due to the four factors was seen in the speed measures (Time, Log_e(Time)) rather than accuracy. To a large extent, Accuracy remained constant across conditions. It appears that participants took as long as they needed to achieve their ultimate level of accuracy, which is a safe and conservative approach to this test.

	Accuracy	**Time**	**Accuracy/Time**	**Log_e(Time)**
Intervention		F(3, 232) = 6.1 P= 0.001		F(3, 232) = 5.9 P= 0.001
Area		F(2, 232) = 13.9 p< 0.001	F(2,232) = 13.9 p< 0.001	F(, 232) = 14.9 p< 0.001
Workcard		F(1, 232) = 6.2 P= 0.014		F(1, 232) = 7.1 P= 0.008
Simplified English				
Reading Level (covariate)	F(1, 232) =22.3 p< 0.001	F(1, 232) = 9.3 P= 0.003	F(1,232) = 18.7 p< 0.001	F(1, 232) = 7.5 P= 0.007
Age (covariate)	F(1, 232) =17.4 p< 0.001	F(1, 232) = 11.7 P= 0.001	F(1,232) = 17.1 p< 0.001	F(1, 232) = 9.7 P= 0.002

Table 3. Summary of ANOVA results for intervention performance

Interventions were only different on the Time measures, although their Accuracy/Time measure approached significance at p = 0.069. The mean times and accuracies for the four interventions are given in Table 4. *Post hoc* Tukey tests showed that for times only the slowest (No intervention) and the fastest (Chinese translation) differed significantly at $p < 0.05$.

Intervention	**Mean Accuracy, percent**	**Mean Time, s**	**Accuracy / Time (%/s)**
1. No Intervention	73.2	1638	4.9
2. Chinese Translation	72.0	1367	5.6
3. Bilingual Glossary	73.8	1469	5.6
4. Bilingual Coach	78.2	1437	5.9

Table 4. Performance Results for the four interventions. Shaded results not significant at $p < 0.05$.

The three areas also differed on Time, but also Accuracy/Time. *Post hoc* Tukey tests at $p < 0.05$ showed that for both measures, the best performing area (Hong Kong) was significantly different from the other two (China, Taiwan). Data for the 15 participants from the USA and UK collected in 2003 are included for comparison, although no statistical tests were performed. These participants were much faster than our Asian sample, but less accurate, resulting in slightly higher Accuracy/Time performance score.

Area	**Mean Accuracy, percent**	**Mean Time, s**	**Accuracy / Time**
1. China	73.0	1519	5.3
2. Hong Kong	78.2	1128*	7.3*
3. Taiwan	75.9	1506	5.5
UK/USA 15 participants	65.8	924	8.0

Table 5. Performance comparisons by Area. Shaded results not significant at $p < 0.05$. Note that * signifies a mean value different from the others at $p < 0.05$ on the *post hoc* Tukey test.

Finally, the results for the two workcards were only significant for the two speed measures, with the easy workcard being faster than the difficult one as expected. Note that this did not happen in our pre-test with Chinese graduate students at an American university.

Workcard	Mean Accuracy, percent	Mean Time, s	Accuracy / Time
1. Easy	73.5	1373	5.9
2. Difficult	74.7	1580	5.2

Table 6. Performance comparisons between the two workcards. Shaded results not significant at $p < 0.05$.

3.3.1 Rating scales

Identical GLM ANOVAs were performed on the fourteen rating scale values, i.e. using Reading Level and Age as covariates and Area, Workcard, Simplified English and Intervention as factors. The major pattern to the results was that on 11 of the fifteen scales, Intervention was the only significant factor, with $p<0.001$ in all of these cases. For 8 of these 11 scales, the only difference in post hoc Tukey tests at $p = 0.05$ was between translation and non-translation. One of the 11 scales showed no significant contrasts while the remaining two only found translation different from the Glossary condition. In all cases, the Chinese translation was rated worse than the other interventions, perhaps reflection the participants' concerns for accuracy of translation from original English documents. Figure 3 compares the mean scale ratings of translation and non-translation interventions for all 15 scales, whether significant (11 scales) or not (4 scales) as noted in the caption.

The other significant results for rating scales were few:

- Workcard × Simplified English for rating scale 2 (Continuity of information) $p = 0.023$
- Area for rating scale 4 (Chance of missing information) $p < 0.001$, and also for rating scale 12 (Compatibility with supplementary information) $p < 0.001$. For both, the only difference was that Taiwan rated significantly lower than the other two areas.
- Age was a significant covariate for Rating Scale 13 (Amount of graphical information)

3.4 Focus Groups

The major characteristics of each site can be found in Appendix Table 1. Supplementing this information, a native English-speaking moderator and an English/Chinese bilingual assistant facilitated the 11 focus groups held across all sites. Focus groups were encouraged to use the language they felt most comfortable with during the discussion: Chinese, English, or both. Each session lasted about 30-45 minutes and was audio taped. The main points are summarized below from combined notes and transcripts.

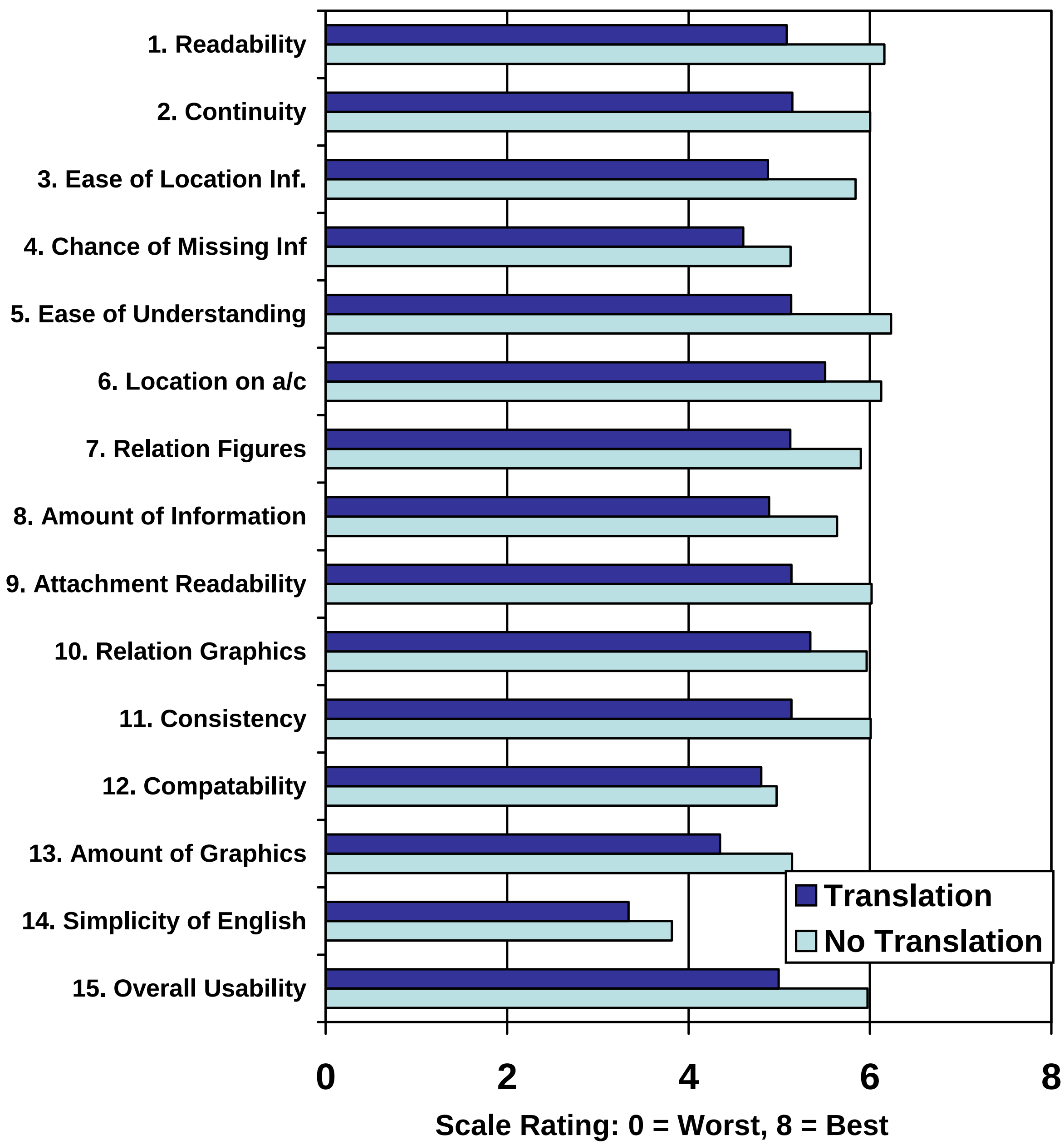

Figure 3. Differences between Chinese Translation and the average of all non-translation groups on the fifteen rating scales.

3.4.1 Current Practice

1.Written Communication to AMTs.

- In China, workcards are bilingual, while Non-Routine Repair forms (NRRs) can be either Chinese or English, [however] companies encourage English. In contrast, workcards and NRRs are in English for both Hong Kong and Taiwan.
- Maintenance manuals are in English for all three areas. The focus groups agreed that Aircraft Maintenance Manual's English is relatively simple (e.g., simple grammar and sentence structure, short sentences).
- Focus groups complained about difficulty in comprehending aspects of English:
    - Long sentences in FAR/JARs, especially those documents related to legal interpretations. In this case, even the Chinese translation is difficult to understand. Even the original English documents can be ambiguous, which results in misunderstanding. In particular, English originals may not be detailed enough, with many steps omitted by the editors of workcards.
    - Multiple meanings for an English word (especially abbreviations). Abbreviation is a general problem: Many questions about English meanings are questions about abbreviations. Different manufacturers use different words or phrases to describe the same thing in their manuals.
- The same English words may be translated into different words in China, Hong Kong, and Taiwan.

2. Writing communication from AMTs

- The shift handover document is often in English, or in Chinese with technical words left in English.
- NRRs are written in Chinese by the mechanics, and then translated by engineer and manager because the international customers require English to be used in NRRs.
- There is a distinct Chinese style of English that can be understood by fellow Chinese colleagues but not by non-Chinese colleagues and manufacturer's representatives. Considerable management effort is spent on rewriting English written by the employees.

3. Verbal communication:

- In China:
    - Most engineers, QA personnel, and leads/foremen speak English. At one site in China, three languages were used in the production meeting every morning.
    - The technicians' oral English ability is often poor.
    - The level of English and Technical English is good in the young generation of aircraft maintenance trainees, but their oral English is still poor.
- At sites in Hong Kong, there is a barrier between the "Mandarin" and "Cantonese" sub-languages e.g., at the maintenance control center.
- Local "nick names" are used in daily work, e.g., "turtle shell." Everybody in the

shop knows what it is, but other people do not necessarily know, and cannot write it down.

4. Company policy
   - Focus groups agreed that English is aviation language, which is a fact that nobody can change. There should not be localized variations. All personnel should emphasize reinforcing standardization inside the company. The focus groups believed that it is perhaps the company's fault that it allows two languages to co-exist at work place. MROs must be more aware of international standards to stay competitive, e.g., comparing themselves with other maintenance bases that only use English workcards.
   - In reality, MROs will continue work with manufacturers that use very different English: "French" English, "Brazilian" English, and American English. These manufacturers may use very different names for the same thing. MROs will also continue work with manufacture representatives who have different language backgrounds (e.g., American English vs. British English).
   - New technology has brought changes to the aviation vocabulary. New words cannot be found in the dictionary, even onsite manufacturer representatives were not sure about them.

3.4.2 Intervention Methods

1. Better Design of Documentation.
   - Translation of workcards was the option used at the sites in China, including initial translation, auditing, and second auditing. Feedback forms to report problems identified onsite with the translated workcards were available, as they are in all organizations using workcards. Focus groups agreed that translation might not be the ideal solution, because:
     - Translations are currently done by college Chinese graduates who are English majors. They have relatively shallow comprehension of aircraft, and short (or even no) working experience on the aircraft. The translation is often based on the obvious non-aviation meaning of the English without in the context of the aircraft. The mechanics find the translations can be confusing, awkward, and even strange.
     - The sheer amount of translation/auditing involves expenses of staff plus an overhead loss and about 30-40 % of the total maintenance time. These all increase maintenance services cost and make the MROs less competitive.
     - The translation/technical writing/editing group has many personnel, each with their own styles in choosing words and structure sentences, e.g., calling a part several different names, which can be confusing to the Chinese mechanics.
   - Provision of both English and Chinese translated versions can help but it is not the final solution—simply because it is impossible to translate everything. For example, there are frequent modifications from the manufacturers.

    - Sometimes reading originals in English is easier than using Chinese translation, especially where the Chinese meaning and English meaning sometimes do not match very well. Occasionally, translations make it worse, e.g., "on/off" "close/open" can be translated into exactly the same Chinese words.
    - Translations of technical references, operation procedures/materials/tools cannot be exactly perfect. There must be English originals available.
- Most MROs have dedicated special focused effort (e.g., company training center, language committee) to develop language references such as:
    - Abbreviation/acronym dictionaries.
    - Glossaries, which were developed by "data mining" for most used Chinese words in the Maintenance Manual.
    - A "Pocket book" consisting of a Chinese-English/English-Chinese dictionary. Most mechanics carry a well-worn copy of this pocket book.
- Focus groups demanded that original English documents use standardization and Simplified English in order to:
    - Be able to use translation software.
    - Clarify the confusion caused by non-native English speakers' lack backgrounds of words.
    - This is especially true for regulations, technical stuff, e.g., get rid of the double negative, which can be confusing rather than emphasizing.
- Mechanics do appreciate diagrams. - Increase numbers of illustrations and diagrams (especially emphasize different angles and positions).
- All would prefer manufactures to provide reference links in its maintenance manual CD-ROMs to reference other documents.

2. Better Education, Training and Language assistance:

- English ability criteria have been used to hire and evaluate performance. English classes have become a part of the curriculum to train apprentices. Some technical classes will be taught in English in the near future. Apprentices are required to pass specific English tests to graduate, and more tests to become certified or promoted. Performance evaluation should always include English. Require certification of English ability integrating with technical/management types of certifications. A small number of employees are selected to study English in local universities every year.
- Engineers are typically on call 24/7 for help with English on a project-by-project basis. People are good at going to supporting engineers for help. However, an engineer often works with many mechanics on the same shift. Mechanics consult engineers for trouble shooting, e.g., checking the Chinese translation and English originals.

Finally, the focus groups have confirmed that there are incidents caused by language barriers. Some examples are:

Case #1: One MRO had an incident caused by "language" resulting in engine damage in 2001. The English word "Clean" has two meanings: 1) get rid of paint, e.g strip, and 2) use cleaner to clean. The correct interpretation should be "get rid of paint; strip" in this context. However, the mechanic did not understand, and performed cleaning by "use cleaner to clean," which resulted in wires being burned from the cleaning fluid.

Case #2: On a test procedure in a manual the Chinese translation did not correctly point out that the "115-160 voltage" should be adjusted continuously rather than being switched. Damage to the aircraft resulted.

Case #3: Slipping Ladder for emergency door: Different people have written the descriptions of the emergency door in different places in the Maintenance Manual. They used different words in different places of the workcard to mean the same thing. The mechanics could not tie the safe-wire the way it was illustrated in the manual. In the end, they had to discuss the problem with the manufacturer and follow their faxed instructions and illustrations. Due to time difference and language barriers, the discussion lasted over 3 days, which prolonged the maintenance process.

4.0 DISCUSSION

For the reasons stated in the Introduction, Asia was the appropriate first region to collect language error data: Asia is a growing center for MROs and the manufacturer's survey reported earlier showed that English usage was relatively low, with translation not often used as an intervention. Our sample of 254 participants across nine MRO sites in three areas (China, Hong Kong and Taiwan) was certainly adequate for testing the incidence of language errors and the effectiveness of potential interventions.

The overall picture was that sites in China used translation into English for at least some workcards, while the other two areas did not. (The maintenance manuals were not translated at any site.) This area difference in response to the obvious mismatch between the language of the maintenance documents and the language of the workforce made sense when the demographics were compared. China had significantly less English reading ability than the other two areas, perhaps because the maintenance of Western airliners is a much more recent activity there. The workforce reflected this, being younger, less experienced, with fewer years learning English and lower Reading Grade levels. Having said that, written English comprehension was at quite a high level throughout: about 5th grade in China and about 6th grade elsewhere. In England and USA for comparison, Reading Grade levels were very high, about 14, as has been found in earlier studies of AMTs (e.g. Drury and Ma, 2004[1], Drury, Wenner and Kritkausky, 1999[6]). The 5-6 grade levels of English reflect an often-stated aim of documentation to be written for a "6th Grade level", although such a recommendation was never meant to apply specifically to aviation maintenance English.

The seven Scenarios, developed from our analyses of language error databases and focus groups in the USA and UK, were found to be well-supported in Asia. There were differences in reporting these errors across the three Asian areas, perhaps related to differences in willingness to report any errors to an FAA-sponsored project run by Westerners. Overall, we were delighted with the level of cooperation at all the sites visited, but even in the USA there is a natural reluctance among AMTs to report evidence that could (despite our assurances) be perceived as jeopardizing their license and hence livelihood.

The most frequently reported scenarios were the ones associated with direct communication surrounding the work itself. All four of these had reported return frequencies between 4 and 5 times per year, and reflected imperfect written communication (work documents) or imperfect verbal communication. The written communication difficulties occurred between the user and English documentation, or with imperfect translation of a source document from another (non-Chinese) language into English. As in the UK/USA sample, participants had example of poor English wording from US manufacturers as well as those from European and South American sources. The examples of scenarios collected from our focus groups confirmed this.

Factors seen as influencing scenario incidence had a large measure of agreement across areas. The typical picture for a language error-prone activity is one with:

Complex task and complex instructions
Poorly designed document
Users with low ability in English and low familiarity with the task to be performed
Time pressure to complete the task

When listed in this way, language errors appear to have all of the usual human factors ingredients for error, not just language error. All of these, apart from low ability in English, can be found in standard texts in human factors, such as Wickens and Hollands (2000[11]) as well as those specifically directed at aviation or aviation maintenance (e.g. Garland, Wise and Hopkins, 1999[12]; Reason and Hobbs, 2004[13]). The implication is that if the "usual" error-shaping factors are present, then the "usual" interventions should be effective, e.g. training (Taylor 1993[14]), documentation design (Drury and Sarac, 1997[15]), organization design (Taylor and Felten, 1993[16]; Reason 1997[17]). We see more evidence for effective interventions as we add the results from the intervention effectiveness experiment and the focus groups.

Direct measurement of intervention effectiveness produced significant results, largely consistent across interventions, areas and workcards, i.e. interactions were almost completely absent, making interpretation simpler. First, as expected, Reading Grade level and Age were highly significant covariates across all measures. Younger participants and those with better reading skills performed better, as has been seen in other studies of document comprehension (Chervak and Drury, 2003[8] Drury, Wenner and Kritkausky, 2000[18]). Such results now extend to a non-native English speaking population.

For the main factors in the experiment, the major finding was that participants opted for constant (and high) accuracy, letting speed suffer. That is exactly the response the traveling public and regulators would like to see. In fact, the accuracy in the test performed was about 74% in our Asian sample compared to about 65% in our UK/USA sample, while the times reflected a one-third decrease for the UK/USA sample. The other surprising finding was that the Simplified English intervention has no effect at all: it made performance no better and no worse, in contrast to our earlier findings that Simplified English was most effective for non-native English speakers (Chervak and Drury, 2003[8]). That finding was for non-native English speakers in the USA, so perhaps SE is less useful when applied in a setting where the native language is other than English. It could of course be a result specific to Chinese language speakers, and only future data collection in other countries such as Latin America could refute that assertion.

The main intervention factor showed that only direct translation into Chinese has an effect on performance, specifically on time taken. This was consistent across workcards and geographical areas, with about a 10% time advantage for translation. However, the translation intervention fared worst of all in the rating scale analyses. We shall consider the whole data set surrounding translation as we incorporate results of the focus groups below.

Results from the focus groups covered much ground in wide-ranging discussions, but as our main concern is in effective intervention strategies, we will consider mainly this aspect and integrate results from the other data collection instruments to produce a more comprehensive picture. As an aid to integration, we will use ICAO's SHELL classification of factors affecting human performance: Software, Hardware, Environment, Liveware (individual) and Liveware (social):

Software: This includes both the task itself and the software such as documentation needed to complete the task. Task complexity was a significant factor in the effectiveness evaluation, with the easy workcard being completed about 15% faster than the more difficult one, although with comparable accuracy as noted earlier. Task complexity was also seen as a major contributing factor in the incidence questionnaire. Unfortunately, the required tasks in maintenance and inspection are often complex *per se*, but for future aircraft, any help designing inherently less complex tasks would help reduce errors, including language errors.

Documentation is the main issue in language error, both documentation read by the AMT carrying out the work and that generated by the AMT to report progress and completion. Much focus group discussion was on the documentation issue, the complexity and consistency of documentation were factors recognized in the incidence questionnaire, and one documentation intervention (translation) produced a significant improvement in the experimental evaluation, albeit with negative comments. These focus group discussions went beyond the specific issue of workcards, although these are a vital part of any maintenance task. There were issues of wording of source documents, such as maintenance manuals and even the FAR/JARs. After at least 10 years of data on the error-reduction benefits of better documentation (Patel et al, 1994[7]) there is really no excuse for continuing to produce poor source documents. They, like all other job aids, must be designed for the user (AMT) rather than for the convenience of the producer or the dictates of lawyers. There are even design aids validated for error-reduction effectiveness, such as the DDA (http://hfskyway.faa.gov) to help make the research findings more accessible to busy document writers. Specifically, designers need to use a single word for each concept, provide abbreviation support, use simple sentences and lay out work documentation instructions in an easy-to-follow format. Where the procedure branches, e.g. as a result of an inspection, a flow chart of the procedure is helpful. Within the body of the workcard, logical branches should follow a standard format, e.g.

```
IF (condition)
    THEN (procedure step)
    ELSE (alternate procedure step)
```

Many of the above factors are part of Simplified English, which did not prove significant in our experiment. Note, however, that both versions of our documents were well-designed compared with earlier workcards such as those used in the Patel et al (1994[7]) and Chervak and Drury (2000[8]) studies.

All documentation needs to be verified for technical accuracy AND validated by having a representative ultimate user (AMT) perform the task exactly from the instructions. This validation must be by a person outside the documentation design team: just using a document writer with an A&P license is not a validation as such a person knows the task and original engineering documentation too well to act as a naïve user.

Translation is the intervention with the largest potential, both positive and negative. Those sites that used translation fully believed in it, although recognizing its limitations in practice. Those sites not using translation had higher levels of English ability in their user populations and saw the errors possible in translation as potential legal traps. Even where it was used, not all documents were translated in the local language, e.g. maintenance manuals, while some were only occasionally translated, e.g. shift turnover forms, NRRs. Translation did improve comprehension, as expected, but the consistently negative ratings of translated workcards reflect the general dislike of this intervention.

To improve translation where it is used, the focus group data suggests that the translation be performed by people who know aviation maintenance AND English, not just professional translators or people with degrees in English. Aviation has special uses for words that also have common meanings beyond aviation, and the difference may not be apparent to people without deep aviation knowledge. Unfortunately, AMTs with excellent translation abilities are rare, and perhaps expensive, but any other solution risks avoidable language error. Whoever is used for translation, consistency is important. The same words must be translated the same way each time, and purely local words should be avoided to ensure that AMTs can move safely between jobs. Standard usage/style manuals should be available and used, as should approved word lists, for example consistent translations from Simplified English. If translation is not used, AMTs need more English language training and practice (see below) but job aids can help. The typical job aids are glossaries and dictionaries, many examples of which were provided to us at the different sites. There is probably a need and market for an aviation maintenance glossary, abbreviation list and dictionary that could be used across sites with the same language. Note however that most languages have variants, e.g. Cantonese *vs.* Mandarin, that need to be accommodated by alternate versions of such a job aid.

Hardware: The primary hardware intervention found was the use of the aircraft (or component) itself to aid understanding of the wording of documents. Seeing and touching the aircraft structure has a solid basis in science and represents a good practice in maintenance. This use of the aircraft also requires that the diagrams in the work documents match the structure itself, as seen from the point of view of the AMT, and also including an orientation sketch. Again, these are good practices already well-documented in the literature (Patel et al 1994[7]). Computer-based work documentation may help here as it can provide support for multiple user levels, e.g. good and weak readers of English, by using hypertext format (Drury, Patel and Prabhu, 2000 [19]) as well as hypertext links between English and the local language.

Environment: Time pressure was recognized as a factor likely to increase language errors, just as it does other errors. This is well known in the maintenance human factors community (e.g. the Dirty Dozen posters produced by Gordon Dupont[20]) but it still occurs. The issue is not whether it exists, as it probably always will in an industry that tries to maintain schedules despite upsets, but whether the effect of time pressure on errors is recognized by those exerting the pressures. Do managers realize the increased error potential when they demand speed, or when they reward those who “get the job done” while turning a blind eye to cutting corners? After 15 years of maintenance Human Factors Engineering, we still need to ensure that performance-oriented managers (and AMTs who often pressure themselves) consistently choose the “accuracy” side of the Speed/Accuracy TradeOff (SATO) (e.g. Drury, 1999[21]).

Liveware (individual): Low English ability, verbal and written, was seen as a causal factor in language error scenarios, a position supported by the significant Reading Grade Level covariate in the intervention study. Most sites recognized this fact and had taken steps to improve English ability of individual AMTs and support personnel. Some MROs had minimum English language entry qualifications, while most had training programs at various levels of ability and tests at each level of promotion. This puts a large burden on the individuals involved and their organizations, but appears to be a necessary cost of the historical decision to use English as the only official language of aviation. More use of consistent practice in written and verbal English can help maintain language skills, e.g. NRRs in English or parts of meeting conducted in English. These reinforcements can help prevent a local patois of "Chinese English" from taking hold in the organization.

Task familiarity was the other individual variable seen as important in reducing language errors. All AMT start as unfamiliar with each task and develop familiarity with training and repetition. Job assignments should be used to ensure that each AMT becomes familiar with each job in a planned manner, typically starting as an extra hand, then working with an experienced AMT as a coach before performing the task alone. In a busy hangar, this may not always be the easiest short-term assignment arrangement, but it ensures increasing capability over time. As more AMTs are available who are familiar with the task, the scheduling task actually becomes easier over time.

Liveware (social): Human/human interaction is a basic part of Human Factors Engineering, and is intimately related to a cooperative social task such as aircraft maintenance. It also related directly to the language errors we found in this study. Some of the interventions noted above, such as planned task assignment or time pressure, are performed through social means. In the incidence study, the most frequent factor in error discovery was that the AMT asked for assistance or clarification – a social act. The second most frequent discovery factor was that the AMT appeared perplexed, again only important if another person notices. From the focus groups came the need for shift turnovers to be understood across shifts whether in English or Chinese, the use of engineers (or leads or managers) who better understand English to act as coaches, and discussions between personnel on English interpretation.

The social interventions derived from these factors consist of a number of logical steps. The first is providing some technical English language backup on all shifts, although the Coaching intervention did not give a significant improvement in the intervention study. Unless everybody understands English perfectly (that is not even true in nominally English speaking countries), the having multiple personnel address any ambiguity is preferable to one AMT going ahead and making a best guess and subsequent error. The second is to provide consistency between the documents used for the work and the documents prepared by the AMT or inspector. Shift turnover logs and NRRs should be in English for consistency, unless the organization is very confident in its translations. This intervention will also help reinforce English in a planned manner and give all personnel practice in writing and reading English. Finally, all personnel need to be taught how to recognize when their co-workers are having difficulty with English. An AMT may not want to ask for help (although AMTs not being willing to expose their lack of English was not seen as a causal factor in the incidence study), but co-workers and supervisors should be sensitive to the signs that understanding may be imperfect. As with any social skill, training and practice can help.

5.0 CONCLUSIONS (also appears as Executive Summary)

This report extended the University at Buffalo/Federal Aviation Administration study of language errors in aviation maintenance to one world region: Asia. Based on earlier studies in the USA and UK, plus analyses of a number of databases, seven scenarios for language error were developed, and the current study assessed the incidence of these scenarios in site visits to 9 sites covering 254 participants in China, Hong Kong and Taiwan. A second study on the same sample was a direct test of the effectiveness of four interventions. Finally, focus groups were used at each site to explore strategies for mitigating language errors. The level of cooperation with the data collection team at all sites was outstanding.

The scenario incidence study confirmed the scenarios as valid and no new ones were added from the focus groups. The typical picture for a language error-prone activity was one with complex task and complex instructions, poorly designed document, users with low ability in English and low familiarity with the task to be performed and with time pressure to complete the task.

The interventions experiment used a baseline condition of English documents, and then added translation (including the test form), a glossary, a bilingual coach, and a combination of these last two conditions. We used two levels of workcard difficulty, each with and without Simplified English. While there were some differences between regions, differences between interventions were consistent across regions. The translation intervention had some effect, although mainly on the times and our performance measure, rather than on accuracy *per se*. If this indeed reflects practice, then maintenance personnel appear to slow down when they find language difficult, rather than making more errors at a constant speed.

Several practical interventions emerged from all three data collection methods. Design of work documentation is the primary way to reduce written language errors. Good design practice will help and translation, if performed carefully, is a viable option. Individual ability of AMTs, inspectors, managers and engineers in written and verbal English communication can be improved by training and controlled practice. The organizational environment should recognize the deleterious effects of time pressure on errors, and also recognize the symptoms of imperfect communication when it occurs. The organization also needs to plan work assignments to allow AMTs to become more familiar with particular tasks, and to use controlled implementation of English in shift turnover documents and non-routine repair documents to provide planned English practice for all personnel.

Our next task is to repeat this experiment in other continents. The current plan is to visit locations in Central and South America and Europe in Spring 2005.

6.0 REFERENCES

1. Drury, C. G. and Ma, J. (2004). Experiments on Language Errors in Aviation Maintenance. *Human Factors and Ergonomics Society 47th Annual Meeting Proceedings*, New Orleans LA, September 20-24, 2004.
2. Drury, C.G. and Ma, J. (2003). *Language Errors in Aviation Maintenance: Year 1 Interim Report*, William J. Hughes Technical Center, Federal Aviation Administration, Grant #2002-G-025.
3. Carver, R. P. (1987). *Technical Manual for the Accuracy Level Test.* REVTAC Publications, Inc.
4. Chervak, S., Drury, C. G., and Ouellette, J. L. (1996). Simplified English for Aircraft Workcards. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting.* 303-307
5. Drury, C.G. and Ma, J. (2003). Do Language Barriers Result in Aviation Maintenance Errors? *Human Factors and Ergonomics Society 47th Annual Meeting Proceedings*, Denver, Colorado, October 13-17, 2003.
6. Drury, C.G., Wenner, C., and Kritkausky, K. (1999). *Development of process to improve work documentation of repair stations.* FAA/Office of Aviation Medicine (AAM-240). National Technical Information Service Springfield, VA.
7. Patel, S., Drury, C. G. and Lofgren, J. (1994). Design of Workcards for Aircraft Inspection. *Applied Ergonomics*, **25(5)**, 283-293.
8. Chervak, S. C. and Drury, C. G. (2003). Effects of job instruction on maintenance task performance. *Occupational Ergonomics*, **3(2),** 121-132.
9. *Overhaul & Maintenance* (2002). MRO in Asia-Pacific Region. Aviation Week's Show News Online: www.aviationnow.com. Online resources: *http://www.awgnet.com/shownews/02asia1/mro09.htm*.
10. Dennis, W. (2002). MRO to Grow in China. *Aviation Week and Space Technology*.
11. Wickens, C. D. and J. G. Hollands (2000). *Engineering Psychology and Human Performance*. NJ, Prentice Hall.
12. Garland, D. J., Wise, J. A. and Hopkin, V. D. (Eds.) (1999). *Handbook of Aviation Human Factors*, Mahwah, NJ, L. Erlbaum Associates, 591-605.
13. Reason, J. and Hobbs, A. (2003). *Managing Maintenance Error, A Practical Guide*. Burlington, VA, Ashgate Publishing.
14. Taylor, J. C. (1993). *The effects of crew resource management (CRM) training in maintenance: an early demonstration of training effects on attitudes and* performance. DOT/FAA/AM-93/5. Human Factors in Aviation Maintenance - Phase Two Progress Report, Alexandria, National Technical Information Service.
15. Drury, C. G. and Sarac, A. (1997). Documentation design aid development. *Human Factors in Aviation Maintenance - Phase Seven, Progress Report, DOT/FAA/AM-97/xx*, National Technical Information Service, Springfield, VA.

16. Taylor, J. C. and Felten, D. F. (1993). *Performance by Design*. NJ, Prentice Hall.
17. Reason, J. (1997). Approaches to controlling maintenance error. *Proceedings of the FAA/AAM 11th Meeting on Human Factors Issues in Aviation Maintenance and Inspection*, San Diego, CA.
18. Drury, C. G., Wenner, C. and Kritkausky, K. (2000). Information design issues in repair stations, *Proc. Tenth International Symposium on Aviation Psychology*, Columbus, OH, May 3-6, 1999.
19. Drury, C. G., Patel, S. and Prabhu (2000). Relative advantage of portable computer-based workcards for aircraft inspection. *International Journal of Industrial Ergonomics*, **26**, 163-176.
20. Dupont, G. "The Dirty Dozen"
21. Drury, C. G. (1999). Managing the Speed-Accuracy Trade-Off. In W. Karwowski and W. Marras (eds), *The Occupational Ergonomics Handbook*, NY, CRC Press, 677-692.

List of Acronyms

AECMA………………………Aircraft European Contractors Manufacturers Association
AIAA……………………………...American Institute of Aeronautics and Astronautics
AMT…………………………………………………Aviation Maintenance Technician
ANCOVA……………………………………………………Analyses of Covariance
ANOVA……………………………………………………...Analysis of Variance
ASRS……………………………………….. NASA Aviation Safety Reporting System
CACC……………………………………….Civil Aviation Administration of China
CAUC…………………………………………....Civil Aviation University of China
FAA……………………………………………….. Federal Aviation Administration
FAR…………………………………………………..Federal Aviation Regulation
GRE…………………………………………………...Graduate Record Examination
GLM……………………………………………………..General Linear Models
JAA………………………………………….…………Joint Aviation Authorities
JAR…………………………………………………………..Joint Aviation Repair
MRO…………………………………………….……Maintenance, Repair & Overhaul
NASA…………………………………National Aeronautics and Space Administration
NRR………………………………………...……………….Non-routine Repair
NTSB…………………………………………National Transportation Safety Board
OEM…………………………………………………Original Equipment Manufacture
PA………………………………………………………………..Public Address
QA…………………………………………...…………………Quality Assurance
SHELL…………………………..Software, Hardware, Environment, Liveware, Liveware
TOEFL……………………………………………Test of English as a Foreign Language

Appendix Table 1. Background information on the MROs

Area	Site #	Type	External Work/Major Customers	Certification	Number of Employees	Style of Using Workcard in Maintenance
China	1	-A large aircraft maintenance engineering company with large hangar facilities -A full range of airframe, engine and component repairs and overhauls; fully certified for line maintenance services for domestic and international carriers.	More than 40 domestic airlines and more than 20 international carriers.	CAAC, FAA, JAA, ISO 9002	More than 3,500 professional technicians	English-Chinese
China	2	-One of the three largest maintenance bases of Chinese civil aviation maintenance system. -Two large hangers facilities and workshops in one location, with another under construction in a new location. -Maintenance services of Airbus aircraft as well as Boeing that include: D-check performance, avionics repair service, calibration center, engine shop, painting, sheet metal, components & accessory repair, cabin refurbishment, etc.	-18 domestic airlines -22 foreign airlines -Maintenance services for various special planes, chartered planes and general aviation	JAA, ISO9001	1,520	English-Chinese
Hong Kong	3	-An Aircraft service company with maintenance operations & cabin services, technical store, and ground support equipment. -Maintenance of aircraft types include Boeing and Airbus -Maintenance services include: line maintenance, cabin services, ground support,	-Providing technical and non technical services to over 20 customers, which include Chinese airlines and foreign airlines	FAA, UKCAA, HKCAD, CAAC, KCAB, PRATO	632	English
Hong Kong	4	-The largest provider of airframe MRO services in the Asia-Pacific region. -The base maintenance facility is amongst the most technologically advanced in Asia comprising a three-bay hangar plus open maintenance apron -Maintenance services include line maintenance, base maintenance, component overhaul, and engine overhaul.	Over 50 customer airlines worldwide.	HKCAD, JAA, FAA, JCAB, and approximately 15 other national regulatory authorities.	2,500	English

China	5	-The aircraft maintenance base of a major Chinese airline, which is a combination of a former aircraft maintenance factory and an aircraft maintenance company -A combination of repair, inspection, production and machining with large hangarage facilities -Primary maintenance includes aircraft line maintenance, aircraft fuselage maintenance and modification, powerplant maintenance, accessory repair, and inspection and examination.	-Chinese airlines, Chinese air force and navy and other Chinese aircraft engineering service/ maintenance companies	ISO9001:2000	1,993	English-Chinese
China	6	-A second-layer financially independent organization under an airline, which had grown from the former aircraft maintenance station. -A single large hangar. -Scheduled maintenance on aircraft 154, I-86, B737-300, B757-200, ATR72-210A, line maintenance, aircraft components/accessories, and special operations (nondestructive inspection, composite material maintenance and repair, etc.)	-Mainly providing maintenance engineering of aircraft, components and accessories as well as spares provisions to its mother airlines. -Providing contracted maintenance services for other airlines.	ISO9000, CAAC	769	English-Chinese
Taiwan	7	-A maintenance station affiliated with a major Taiwanese airline. -Single large Hangar -Performing aircraft, engine and component repair.	Domestic Airlines, and aircraft leasing services	FAA, ISO9002, CCAA, CAAC, RIDCA, PICAA	About 300 maintenance personnel	English
Taiwan	8	-A joint venture between the Engineering & Maintenance Division of a major Taiwanese airlines and an engine manufacturer -Two hangars, two engine shops and a test cell. Additional hangar under construction in 2004 -Primary maintenance includes component maintenance, engine maintenance, heavy maintenance, ramp maintenance.	-Taiwanese airlines and 12 Foreign airlines	FAA, JAA, CCAA(Taiwan), CAAC (China), AACM, DGCA, ISO-9002, CASE	1,300	English
Taiwan	9	-A maintenance facility affiliated with a major Taiwanese airline. -Single hangar plus component shops	-Maintain only this airlines regional aircraft.	CAA, ACAB, ISO9002, ISO9001-2000	?	English

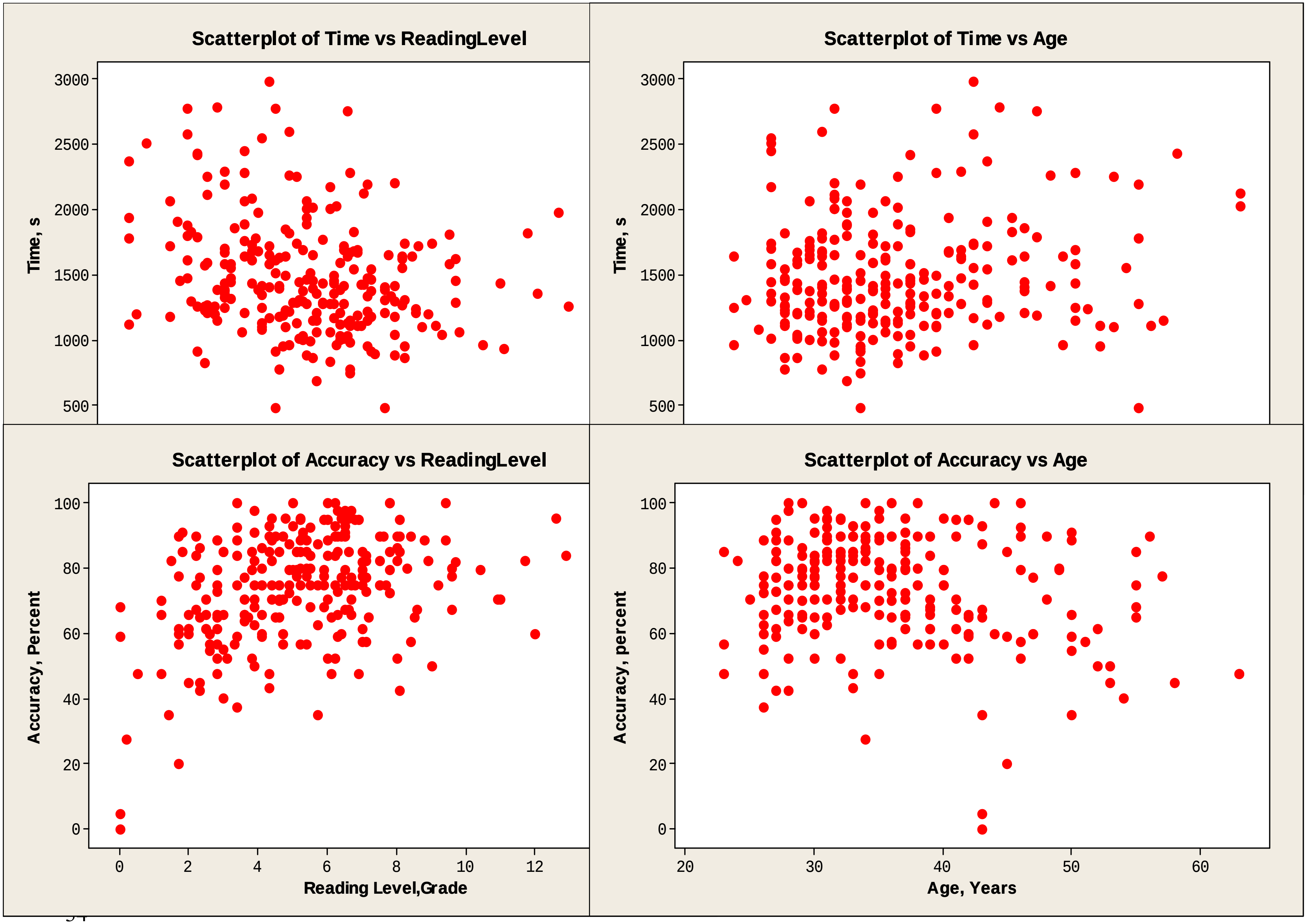

Figure 2. Scatter plots of the two aspects of performance (Accuracy, Time) against the two covariates (Reading Level, Age).